



EUROPEAN PATENT APPLICATION

(43) Date of publication:
04.01.2023 Bulletin 2023/01

(51) International Patent Classification (IPC):
A61B 7/00 (2006.01)

(21) Application number: **21183483.3**

(52) Cooperative Patent Classification (CPC):
A61B 7/003

(22) Date of filing: **02.07.2021**

(84) Designated Contracting States:
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB
GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO
PL PT RO RS SE SI SK SM TR**
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

- **DA CONCEIÇÃO BARATA, Vitor Filipe**
8047 Zürich (CH)
- **OTT, Kai Robin**
3013 Bern (CH)
- **TINSCHERT, Peter Jan**
8006 Zürich (CH)

(71) Applicant: **CSS Versicherung AG**
6002 Luzern (CH)

(74) Representative: **KATZAROV S.A.**
Geneva Business Center
12 Avenue des Morgines
1213 Petit-Lancy (CH)

(72) Inventors:
• **SHIH, Chen-Hsuan**
8003 Zürich (CH)

(54) **METHOD FOR MONITORING ACOUSTIC EVENTS IN A PATIENT**

(57) The present relates to methods for detecting a patient's acoustic events, for instance coughs, during an analyzing period from audio data of the patient, the method comprising:

- a recording step to record an audio signal from the patient by using a mobile device comprising recording means, and
- a processing step to process the recorded audio signal into audio data by using a processing module embedded in the device,
- a detection step to detect said acoustic events from the audio data, characterized in that the method further comprises a monitoring step, before the recording step, to

select segments of a determined duration of the patient's audio signal to be recorded on the device, said monitoring step comprising:

- applying an adaptive loudness threshold named ALT on the patient's audio signal to identify sounds that exceed said adaptive loudness threshold named relevant sounds,
- said adaptive loudness threshold being based on an initial loudness threshold that is configured to be adjusted continuously during said analyzing period, and
- recording on the device the segments of patient's audio signal located downstream said identified relevant sounds.

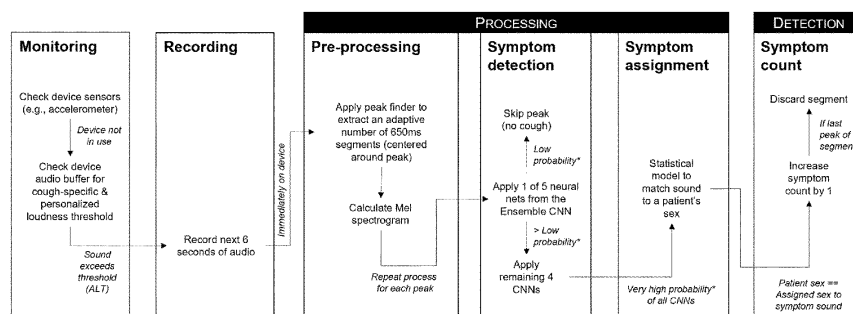


Figure 1

Description**Technical Field**

- 5 **[0001]** The present invention relates to a computer implemented method for detecting a patient's acoustic events, for instance coughs, from audio data of the patient, and to a device comprising means for carrying out said method.
- [0002]** The invention further relates to a computer program comprising instructions, which, when the program is executed by a computer, cause the computer to carry out said.

10 **Background of the art**

- [0003]** Asthma is one of the most prevalent chronic respiratory diseases. Despite increased investment in treatment, little progress has been made in the early recognition and treatment of asthma exacerbations over the last decade.
- 15 **[0004]** Nocturnal cough monitoring may provide an opportunity to identify patients at risk for imminent exacerbations. Recently developed approaches enable smartphone-based cough monitoring by using the microphone of the smartphone to record the cough.
- [0005]** Subsequently the audio file is processed to detect cough event, notably for eliminating the noise or other non-relevant audio events, and to count the cough event for instance overnight.
- 20 **[0006]** An example of approach is described in the document Barata et al, entitled "Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study", J.Med. Internet Res, vol.22, issue 7, 2020.
- [0007]** However, this approach is not practical for real-time analysis on mobile devices, notably because it requires more computational power than is present in mobile devices, for instance it requires the computational power of a supercomputer, to process the audio files recorded by the mobile device.
- 25 **[0008]** Additionally, it requires a lot of storage on the smartphone to record the entire nights of audio file. After recording, a considerable amount of raw audio data must be uploaded to a server (e.g., the Euler supercomputer), which raises data privacy concerns and requires patients to have a fast and reliable internet connection.
- [0009]** Eventually, patients have to wait for results until data upload and analysis on a server to be completed which generates multiple days of delay between recording and receiving results.
- 30 **[0010]** Therefore, there is a need to provide a solution to facilitate the monitoring of acoustic event and overcome the issues of the existing approaches.

Summary of the invention

- 35 **[0011]** The above problems are solved or at least minimized by the present invention.
- [0012]** This object has been achieved by providing a computer implemented method for detecting a patient's acoustic events, for instance coughs, from audio data of the patient during an analyzing period, the method comprising:
- a recording step to record an audio signal from the patient by using a mobile device comprising recording means, and
 - 40 - a processing step to process the recorded audio signal into audio data by using a processing module embedded in the device,
 - a detection step to detect said acoustic events from the audio data,
- 45 characterized in that the method further comprises a monitoring step, before the recording step, to select segments of a determined duration of the patient's audio signal to be recorded on the device, said monitoring step comprising:
- applying an adaptive loudness threshold named ALT on the patient's audio signal to identify sounds that exceed said adaptive loudness threshold named relevant sounds,
 - 50 - said adaptive loudness threshold being based on an initial loudness threshold that is configured to be adjusted continuously during said analysing period, and
 - recording on the device the segments of patient's audio signal located downstream said identified relevant sounds.
- 55 **[0013]** Another object of the present invention is to provide a computer implemented method for assigning a detected acoustic event, for instance coughs, to a patient comprised in a group of patients, the method comprising:
- Providing a recorded segment of audio signal of one of the patients from the group;

- Detecting an acoustic event in each of the patients from the group by using a method according to the present invention;
- Comparing the detected acoustic event with the provided recorded segment of audio signal;
- If there is a match between the detected acoustic event and the provided recorded segment audio signal, assigning the detected acoustic event to said patient.

[0014] A further object of the present invention is to provide device comprising means for carrying out the method of the present invention, preferably the device is chosen among a mobile device, for instance a smartphone.

[0015] A further object of the present invention is to provide a computer program comprising instructions, which, when the program is executed by a computer, cause the computer to carry out the method of the present invention.

Description of the invention

[0016] The invention concerns a computer-implemented method for detecting a patient's acoustic events, for instance coughs, from audio data of the patient during an analyzing period, the method comprising:

- a recording step to record an audio signal from the patient by using a mobile device comprising recording means, and
- a processing step to process the recorded audio signal into audio data by using a processing module embedded in the device,
- a detection step to detect said acoustic events from the audio data,

characterized in that the method further comprises a monitoring step, before the recording step, to select segments of a determined duration of the patient's audio signal to be recorded on the device, said monitoring step comprising:

- applying an adaptive loudness threshold named ALT on the patient's audio signal to identify sounds that exceed said adaptative loudness threshold named relevant sounds,

said adaptative loudness threshold being based on an initial loudness threshold that is configured to be adjusted continuously during said analysing period, and

recording on the device the segments of patient's audio signal located downstream said identified relevant sounds.

[0017] The logic behind the present invention is that patients tend to cough (example of acoustic event) rather infrequently during a period of time, for instance during the night. Therefore, it is possible to vastly limit the number of computations needed to detect an acoustic by excluding or discarding as many parts of the night for instance as possible in which no acoustic event occurs (i.e., patients most likely do not cough). If the method according to the present invention is executed only at times when an acoustic event does occur (when patients cough), namely for relevant sounds, a vast majority of computations can be spared without scarifying the accuracy of the detection since the most relevant part of the audio signal will be analysed. This approach allows to limit the computation power required to execute the detection process.

[0018] Advantageously, the method, for instance, the computer program or software related thereof, can be run on a smartphone since computation and analysis are limited to the most relevant part of the signal, namely relevant sounds. In other words, all the necessary calculations can be performed by using exclusively the computational power of a smartphone.

[0019] The present invention also allows to limit the storage capacity required on the mobile device executing the computer implemented method since it does not require a continuous recording of the audio signal or audio data.

[0020] In general, if the threshold is high, fewer acoustic events will be recorded and analyzed, but a relatively higher number of acoustic events could not be detected if they are too silent. For a low threshold a higher number of coughs can be detected, but the computational resources (CPU, memory, power) usage will increase, since more acoustic events need to be analyzed. Moreover, the threshold can vary from patient to patient and/or from one analysing period to another (night to night) for various reasons (e.g., different cough loudness, bedroom outline, smartphone placement, background noise). Therefore, setting a static threshold, that is the same for every patient, is not adequate. In the present invention, an adaptive loudness threshold is used. The point of this adaptative loudness threshold is to find the most suitable loudness threshold per patient and per analysing period (per night for instance night).

[0021] Advantageously, the present invention offers an adaptive threshold that aims at finding the ideal loudness threshold per patient, and also per recording period (or recording session) for a same patient (i.e. per night or day).

[0022] In the recording step of the present invention, an adaptative loudness threshold is used to assess whether a segment of the audio signal is relevant, i.e., relevant sound, (likely to contain an acoustic event) or not. An initial loudness threshold is set as a basis or reference loudness threshold. The loudness threshold is adaptative, meaning that the loudness threshold is adjusted continuously to personalize the threshold to the patient and the patient's located environment. In other words, the initial loudness threshold is adjusted depending on the sound level of the acoustic events (i.e., retroactively) in the patient during the recording or while the signal is recorded.

[0023] Advantageously, the adaptative loudness threshold is patient-specific and/or environment-specific, meaning that the loudness threshold is adjusted to respectively the audio signal of each patient and each environment allowing a personalized recording.

[0024] Advantageously, the audio signal having a sound level below the adaptative threshold is not recorded, which allows keeping the storage capacity of the mobile device for relevant sounds. This allows to limit the storage capacity required on the device.

[0025] Additionally, when a relevant sound detected, the recording is initiated to record the portion of the audio file, namely a segment, downstream or immediately after or next to the detected relevant sound. In other words, the detection of a relevant trigger the recording of the segment of a determined duration.

[0026] Contrary to the existing approaches, the present invention avoids processing the audio signal continuously for sound below a loudness threshold but rather aims at processing discrete or chosen portions of the audio file exceeding the adaptative threshold.

[0027] The adaptative loudness threshold permits to record overall less audio signal, which means that less data needs to be analysed and thus computational resources are spared. The audio file selections are made without sacrificing on the accuracy or with a very limited or negligible effect on the accuracy of the acoustic event detection. The reason is that only or mostly irrelevant parts of the audio signal are excluded in which there is no acoustic event or in which the chance to have an acoustic event is very low or negligible.

[0028] Preferably, the analyzing period is variable. For instance, the analyzing period can be chosen by the user patient, who can start and stop the monitoring manually.

[0029] The analyzing period is preferably comprised between about 4 hours and about 14 hours, preferably between about 6 hours and 12 hours, for instance 8 hours or 12 hours.

[0030] The analyzing could be started when the patient goes to sleep and stopped when the patient gets up, for instance on average 8 hours.

[0031] chosen to be at least 1 second and not restricted in length, more preferably between 1 second and 24 hours, for instance 12 hours.

[0032] In preferred embodiment, the adaptative loudness threshold is adjusted based on at least three parameters defined as:

- a first parameter named T and based on the loudness of the detected acoustic events from the patient;
- a second parameter named T_N and based on a factor M times of the average loudness of the audio signal;
- a third parameter named T_{min} and based on a determined minimum threshold;

wherein said adaptative loudness threshold ALT corresponds to the maximum value chosen among the first parameter, the second parameter and the third parameter and is defined as

$$ALT = \max (T, T_{min}, T_N)$$

[0033] In other words, during the analyzing period, the threshold varies and corresponds to one the parameters, namely either to the first parameter, to the second parameter, or to the third parameter.

[0034] For example, if the first parameter is -24.8dB, the second parameter is -38dB and the third parameter is -32dB, the adaptative loudness threshold corresponds to the first parameter, that is -24.8dB. The maximum of negative numbers is the least negative.

[0035] Preferably, the first parameter T is based on the distribution of the acoustic event detected in the patient, said distribution being adjusted continuously as the acoustic events are detected. Thus, the adaptative loudness threshold can adjust to the loudness of the patients' cough, room size, distance to phone, orientation etc.

[0036] Preferably, the factor M time of the second parameter T_N is chosen between 0 and about 100, preferably about 2 and about 50, more preferably between about 4 and about 25, in particular between about 8 and about 12.5, for instance between about 9 and about 11, for instance 10. Thus, in a very noisy environment, the number of acoustic event (cough for instance) discarded or sacrificed is very limited (typically a few) to analyze way fewer audio signals

(that would be mostly caused by loud noise).

[0037] Preferably, the minimum threshold of the third parameter is comprised between about -18dB and about -45dB, preferably between about -22dB and about -42dB, more preferably between about -27dB and about -37dB, in particular between about -30dB and about -34dB, for instance between about -31dB and about -33dB, for instance about -32dB. Thus, if the acoustic events (coughs for instance) that can be monitored or recorded are really silent (maybe the patient is too far from the recording means), the threshold is not lowered indefinitely, which would induce the analysis of silent (weak) audio signals.

[0038] Preferably, the determined duration of the segment is comprised between about 1 second and about 20 seconds, preferably between about 2 seconds and about 15 seconds, more preferably between about 3 seconds and about 10 seconds, in particular between about 4 seconds and about 8 seconds, for instance between about 5 seconds and about 7 seconds, for instance about 6 seconds.

[0039] Advantageously, the duration of the segment is chosen to limit the impact on the storage capacity of the device. Preferably, the duration of a segment is in seconds, is comprised between about 1 second and about 20 seconds, preferably between about 2 seconds and about 15 seconds, more preferably between about 3 seconds and about 10 seconds, in particular between about 4 seconds and about 8 seconds, for instance between about 5 seconds and about 7 seconds, for instance about 6 seconds.

[0040] In a preferred embodiment, the method further comprises, before the monitoring step, a status check step comprising:

- determining if the device is in use or not in use based on at least one sensor of the device, and starting the monitoring step if the device is not in use; and/or
- determining if the device has an energy level above a predetermined level, and starting the monitoring step if the energy level is above the predetermined level;

[0041] The status check filter is particularly adapted for detecting acoustic events when the patient is asleep, for instance during nighttime. The logic of the status check filter is that the status of the mobile device is correlated with the state of the patient. In other words, if the device is in use, the patient is not asleep. Determining the status of the device allows execution of the method only when the patient is asleep.

[0042] The status check filter can also check if the battery level is sufficient to execute the method of the present invention, i.e., above a predetermined energy level. The point is for instance to avoid that the execution of the method of the present invention drain the entire battery and prevent the execution of other actions or feature on the mobile device, for instance, an alarm if the device is a smartphone. For instance, the predetermined energy level is between about 10% and about 40 %.

[0043] Preferably, status check is based on sensors installed in the device capable of giving information on the position of the device, for instance an accelerometer, and/or on the status of the device for instance, screen on/off and brightness, battery status.

[0044] In a preferred embodiment, the processing step comprises a pre-processing step that comprises:

- Applying a peak detection filter on said recorded segments to extract a determined number of peaks;
- Processing the extracted peaks of the patient's audio signal to provide the audio data;

[0045] The peak filter aims at identifying distinct peak(s) or acoustic peak(s) within the recorded segments instead of processing the entire whole audio file or the computationally expensive sliding window logic of existing approaches such as Barata et al (cf para.[006] above). In the present case, only the extracted peaks of the audio file are processed to provide the audio data, for instance audio data as Mel spectrograms.

[0046] This allows saving computational resources without sacrificing the overall accuracy in a meaningful way since the computational resources process the relevant sounds.

[0047] For example, the audio data that can be used in the present invention as a soundwave could be:

- kept in the format it is in,
- or processed to a normal spectrogram,
- or processed to a Mel spectrogram,
- or discretely transformed to Fourier space,

- or converted to an image of the soundwave,
- or transformed with the Discrete Wavelet Transform,

5 **[0048]** The number of peak detectable per recorded segment is adaptive and is determined depending on the processing module specifications, for instance, the processing speed or the calculation power of said processing module, or the acoustic environment. This allows seamless execution of the peak finder depending on the computation power of the device.

10 **[0049]** The peak filter can be optimized depending on some setting of the device, for instance, the availability of the device's computational power and battery. For instance,

- in a device whose processing unit is not fully in use, the number of peaks per 6 seconds can be increased,
- in a device whose processing unit is fully in use, the number of peaks per 6 seconds can be decreased,
- and where the default is between around 3 peaks / 6 seconds and around 12 peaks per 6 seconds, for instance 6 peaks per 6 seconds.

15 **[0050]** In a preferred embodiment a minimal distance is set between each peak, preferably personalized to each patient, that is comprised between about 0.2 seconds and about 1.1 seconds, preferably about 0.2 seconds and about 0.8 seconds, more preferably between about 0.2 seconds and about 0.40 seconds, in particular about 0.35 seconds. This prevents a peak, and hence an acoustic event, from being counted multiple times. The inventors found out that generally only less than 5% of peaks are closer to each other than .35s. This also allows for a higher accuracy of the present invention in comparison to applying a peak finder without this boundary condition.

25 **[0051]** Preferably, the processing step comprises a pre-processing step that comprises:

- Applying at least one preprocessing filter on the recorded segments of audio signal to provide a selection of segments, preferably said preprocessing filter being chosen from the list comprising entropy filter, voice activity detection (VAD), preferably an entropy filter followed by a voice activity detection (VAD);
- Applying a peak finder on said selected segments to extract a determined number of peaks within each selected segment;

30 **[0052]** The preprocessing filter permits to exclude segments without having to compute any audio data, such as Mel spectrogram for them, which is a computationally demanding task. In this way, only segments that have likely relevant acoustic events in them are processed.

35 **[0053]** In a preferred embodiment, the detection step comprises

- i) Applying a first trained inference model on said recorded segment of the audio data of the patient as an input, to compute a probability as an output, said audio data contains at least one acoustic event, said first trained inference model belonging to an ensemble of trained inference models comprising "n" trained inference models, "n" superior to two, to detect said acoustic events; and if the probability is inferior to a first reference value, the acoustic event is rejected, whereas if the probability is superior to the first reference value, step i) is repeated with the next trained inference model from said ensemble comprising "n" trained inference model; and
- ii) the acoustic event is validated if the probability is superior to the first reference value for all the "n" trained models, and if there is a mean probability for all the "n" trained inference model of the ensemble that is above a second reference value;

40 **[0054]** Advantageously, the ensemble of trained inference models is used to detect the acoustic event in the audio data. For instance, said ensemble is used to derive output from the inference models that process the Mel spectrograms.

45 **[0055]** The recorded segment of audio data is used as input to an ensemble of trained inference models to compute a probability and a mean probability as output. Derivative of said recorded segment can also be used as input such as Mel spectrograms. Depending on these computed probabilities, the acoustic event is validated (i.e., counted) or rejected (i.e., not counted).

50 **[0056]** Advantageously, a two steps approach is used herein based on two reference values, namely a first reference

value and a second reference value. Overall, the point is to obtain a mean probability above a second reference value to have an acoustic event validated.

[0057] The mean probability of all the trained inference models depends on each probability of each trained inference model. Therefore, in the first step, one of the trained inference models is executed to compute a probability. The probability computed by the first trained inference model is used to determine if the remaining trained inference model of the ensemble should be executed or not. In other words, the first computed probability computed by the first trained inference model determines if the rest of the trained model shall be executed or not. If the first computed probability is inferior to a first reference value, the mean probability for all the trained inference models cannot be above the second reference value, so that there is no need to execute said remaining trained inference model.

[0058] For instance, let assume the ensemble of trained inference models comprises 5 (five) trained inference models, the first reference value is set to 75 % and the second reference value set to 95%:

- If the probability computed by the first trained inference model is below 75%, the mean probability of 95% cannot be reached any more; there is no need to execute the remaining 4 (four) trained inference models and the audio data corresponding to said peak is not counted as acoustic event.
- If the probability computed by the first trained inference model is above 75%, the mean probability of 95% can be reached; the remaining 4 (four) trained inference models are executed and if the mean probability of all the ensemble of trained inference model is above 95%, the audio data corresponding to said peak is counted as acoustic event.

[0059] The point of these two steps' approaches herein save the computational power without sacrificing any accuracy.

[0060] Preferably, wherein the first reference value is correlated with the mean probability for all the inference models of the ensemble and the number "n" of trained inference models by using the following formula:

$$\text{First reference value} = 1 - ((1 - \text{second reference value}) \times \text{number "n"})$$

[0061] Preferably the trained inference model is chosen among a neural network classifier.

[0062] In a preferred embodiment, the ensemble of trained inference models used in the present invention is described in the document Barata et al, entitled "Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study", J.Med. Internet Res, vol.22, issue 7, 2020.

[0063] Preferably, the first reference value comprised between about 50 % and 100 %, preferably between about 60% and about 90 %, more preferably between about 70% and 80%, for instance about 75 %.

[0064] Preferably, the second reference value is comprised between about 50 % and 100 %, preferably between about 70% and about 100 %, more preferably between about 80% and 100%, more preferably between about 90 % and 100 %, for instance about 95 %.

[0065] Preferably, the patient is comprised in a group of "p" patients comprising at least two patients, the method further comprising an assigning step for assigning a detected acoustic event to a patient from the said group, the assigning step comprising:

- Applying a computational tool, for instance a statistical model on the detected acoustic event, said computational tool being configured to assign a detected acoustic event to a predetermined patient;
- Attributing the detected acoustic event to one patient from the group if there is a correlation between said computation tool and the detected acoustic event;

[0066] The computational tool can be a trained inference model or a machine learning model such as a Neural Network or Random Forest classifiers. The computational tool can also be a statistical model.

[0067] Preferably, the computation tool has been trained to assign a detected acoustic event to a type of patient is based on gender, for instance male or female, the age for instance child or adult. In other words, the predetermined patient is for instance a male or a female, a child or an adult. Advantageously, it is possible to have one computation tool, for instance pre-trained model for each type of predetermined patient so as to assign to a specific type of patient. There is no need to have a sample of previously detected acoustic event for each patient whose audio data is analyzed.

[0068] Alternatively, the computation tool has been trained to assign a detected acoustic event to a designated patient based on previously detected acoustic event from said designated patient. In other words, the predetermined patient is one specific patient for who acoustic events have been already detected. The previously detected event will be used as a template to assign the acoustic event to this specific patient. Advantageously, the assigning step allows to assign

coughs to a patient not based labels like male or female, but by comparing a detected acoustic event with a patient acoustic event that was recorded earlier. Additionally, it is possible to assign acoustic event to patient, even if they are of the same group (for example with a same gender). For instance, when a female patient coughs, it is possible to assign that cough to her even if she shares the room with another female person who coughs.

[0069] The assigning step can be used in a group of patients. It can also be used for one patient. This is advantageous for instance if the gender of the patient is unknown, if the age of the patient is unknown, if the identity of the patient is unknown.

[0070] When an acoustic event is detected, a computation tool (for instance a model or statistical model) will be used to label it or assign it to a predetermined patient or predetermined type of patient, for instance a male or a female. Knowing the gender of the patients from the group of patients, it is possible to know if the acoustic event is of the patient (genders match) or of their partner. Alternatively, a model can label or assign a detected acoustic event as "adult cough" or "child cough" for example, which would help to assign coughs to a patient, if the e.g., adult patient is in the same room as a child who also coughs.

[0071] The particular advantages of the methods are similar to the ones of the device of the invention and will thus not be repeated here.

[0072] The embodiments describe for the computer-implemented method also apply to the computer program and to the device according to the present invention mutatis mutandis.

Brief description of the drawings

[0073] Further particular advantages and features of the invention will become more apparent from the following non-limitative description of at least one embodiment of the invention which will refer to the accompanying figures (tables and plots), wherein

- Figure 1 is a description of the method for detecting a patient's acoustic events, here coughs according to an embodiment of the present invention;
- Figure 2 represents an embodiment to continuously adjust the adaptative loudness threshold according to the present invention;
- Figures 3,4,5 represent the evolution of the adaptative loudness threshold in three examples - respectively example 1, example 2 and example 3 - of the method according to the present invention;

Detailed description of the invention

[0074] The present detailed description is intended to illustrate the invention in a non-limitative manner since any feature of an embodiment may be combined with any other feature of a different embodiment in an advantageous manner.

[0075] In the examples represented in figures 1 to 5, the mobile device used in the present invention is a smartphone and all the calculations required to run the method are performed by using the computation resources of the smartphone. Additionally, in the examples represented in figures 1 to 5, the acoustic event that is detected is cough. However, the present invention is not limited to smartphones as mobile devices or cough as the only detectable acoustic event.

[0076] Figure 1 is a scheme illustrating an embodiment of the method for detecting patient's cough from audio data during an analyzing period of 8 hours during the night (patient asleep) for instance in the present embodiment.

[0077] The method comprises four main steps, namely a monitoring step, a recording step, a processing step and a detection step.

[0078] The first step is the monitoring step which aims at assessing whether a specific segment is relevant. To that end, the device checks the audio buffer continuously for noises that exceed an adaptative loudness threshold (to be discussed here above with figure 2). The monitoring step permits to select segments of a determined duration of the patient's audio signal to be recorded on the device. In the present example, the determined duration is 6 seconds.

[0079] The monitoring steps also comprise a status check step to check continuously (for instance every 10ms) whether the device is currently in use based on other sensors (i.e., accelerometer, screen brightness, and battery status). Only if the device is not in use and has a sufficient battery level, the method (i.e. detection pipeline) is executed. The innovative improvement of this step is two-fold: 1) it intends to analyze nocturnal cough (i.e., cough in the nighttime is defined through the sleep state of a patient, and if the device is in use, the patient is not asleep) and 2) it prevents impairment of patients' user experience when using their device for other purposes than cough detection (if the pipeline is executed it requires computational resources which then lack for the seamless execution of other smartphone applications; if battery state is not considered, the detection pipeline may drain the entire battery, so that an alarm to wake up the patient in the morning would not be executed).

[0080] During the monitoring step, when a sound of the audio data exceeds the adaptative loudness threshold (sound is named a relevant sound), it triggers the recording of the next 6 seconds of audio data on the mobile device in the

recording step. In other words, the detection of a relevant sound induces the recording of the 6 seconds audio signal located downstream of this relevant sound.

[0081] Subsequently, the processing step is executed in the smartphone. In the present invention, the audio signal that is recorded on the mobile device is filtered by the adaptative threshold.

[0082] In the present example, the processing step first comprises a preprocessing step that is constituted by:

- Applying a peak detection filter on the recorded segments to extract a determined number of peaks per recorded segment;
- Processing the extracted peaks of the patient's audio signal to provide the audio data;

[0083] In the present example, the peak detection filter selects up to 6 peaks per 6-second frame. Those peaks are selected in the following manner: From all peaks in the 6 seconds, that is all local maximum points, the peaks with the lowest amplitude will be discarded iteratively until the duration between any two of the remaining peaks is at a distance of at least 125ms. If the number of the remaining peaks is above 6, then the 6 peaks with the highest value will be selected, otherwise, if the number of the remaining peaks is 6 or lower, they will all be selected. The data used for further processing of these up to 6 peaks comprises a 650ms window centered around each peak (that is 325ms to the left/before and 325ms to the right/after).

[0084] Then the extracted peaks are processed as Mel spectrograms.

[0085] The Mel spectrograms are used as input for the symptom's detection steps based on trained inference models.

[0086] In the present example, the trained inference models that are used are described in the document in the document Barata et al, entitled "Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study", J.Med. Internet Res, vol.22, issue 7, 2020.

[0087] The trained inference models described in Barata et. Al, and used in the present example are 5 convolutional neural networks (CNN). These 5 CNNs have been trained on data collected during a longitudinal study. The collected data (sounds) were labeled either as a "cough" or as a "non-cough". Each of the model was trained on the same "cough" data, but on different "non-cough" data. The main features of these models are that they are lightweight and have been optimized for efficiency so that they could be run on processor-wise limited devices like smartphones. For example, these neural networks do not include fully connected layers, which, although common in most neural networks, are computationally expensive.

[0088] Overall, the detection steps based on the trained inference model comprises:

- i) Applying a first trained inference model on the Mel spectrograms to compute a probability as an output, said first trained inference model belonging to an ensemble of five trained inference models in the present example; and
if the probability is inferior to a first reference value, here 90 % in the present example, the cough is rejected, whereas if the probability is superior to 90 %, step i) is repeated with the next trained inference model from said five trained inference model; and
- ii) the cough is validated if the probability is superior to 90 % (first reference value) for all the "five" trained models, and if there is a mean probability for all the "five" trained inference model of the ensemble that is above 98 % (as a second reference value) in the present example;

[0089] Then the detected cough is assigned to either male or female by using a computational tool. In this example, this is done using the 1) output of two Gaussian Mixture Models (GMM) that have been trained on the cough data from the study to identify male and female coughs and 2) using the information provided by the patient (their sex and whether or not they sleep alone) to determine the final output. That is, if the cough is assigned the same sex as the patient sex, then the cough is assigned to the patient. Otherwise, the cough assigned to a third person (partner for instance) and not counted.

[0090] Figure 2 illustrates an example of the process to continuously adjust the adaptative loudness threshold according to the present invention.

[0091] In the present example, the adaptative loudness threshold is based on three components:

- a first parameter named T and based on the loudness of the detected acoustic events from the patient;
- a second parameter named T_N and based on a factor M times of the average loudness of the audio signal;
- a third parameter named T_{min} and based on a determined minimum threshold;

[0092] In the present example, the first parameter based on already detected coughs is determined as follows (the following being only an example of how this first parameter can be set, the invention is not limited to this example of first parameter T). The distribution of cough loudness is estimated as coughs are detected during the analyzing period and from that distribution the threshold T is set.

[0093] The estimated distribution comprises an initial distribution and an empirical distribution calculated from all coughs' loudness, each distribution being weighted and the weight varies during the analyzing period depends on the coughs detected. Starting with the initial distribution, the more coughs are detected, the greater the relative weight of the empirical distribution is.

[0094] The initial distribution itself comprises:

- a static initial distribution, for instance determined empirically from previous studies data, and that is preferably the same for all analyzing period and patients of one sex ; and
- a personal initial distribution, that is based on the estimated distributions of the preceding nights (analyzing period) for the patient

[0095] The initial distribution is a weighted average of the static distribution and the personal distribution.

[0096] The weight of the preceding nights' distribution can be 0 (zero) if not data from the preceding night (analyzing period) is available. Preferably, the weight will depend on the number of nights (analyzing period) the estimation has successfully taken place.

[0097] In the present example, the first parameter T is calculated by using the following formulas:

Symbol	Description	Symbol	Description
T	Threshold based on cough loudness	μ_0	Mean prior (from sample analysis)
μ	Estimated mean	σ_0	Std prior (from sample analysis)
σ	Estimated std	$\hat{\mu}$	Empirical mean
L	Number of detected coughs	$\hat{\sigma}$	Empirical std
$y^{(l)}$	Amplitudes of detected coughs	$\hat{V}$	Empirical variance
$T = f(\mu, \sigma) = \mu - c_0 \cdot \sigma$ where $c_0 = 2.3$ is a tunable parameter			

The empirical values-are calculated during the analysis period as:

$$\hat{\mu} = \frac{1}{L} \sum_{i=1}^L y^{(i)} \quad \text{and} \quad \hat{V} = \frac{1}{L} \sum_{i=1}^L (y^{(i)})^2 - \hat{\mu}^2$$

The estimated mean and variance will then be a weighted average of empirical values and prior:

$$\mu = a \cdot \hat{\mu} + b \cdot \mu_0 \quad \text{and} \quad \sigma = \sqrt{a \cdot \hat{V} + b \cdot \sigma_0^2}$$

with $a + b = 1$ The estimated values tend away from the priors to the empirical ones as the number of samples increases in the following manner:

$a = 1 - e^{-rL}$ and $b = e^{-rL}$ where $r = 0.1$ is a tunable parameter

[0098] In the present example, the second parameter T_N is based on the noise level said noise level being the average loudness in at a time t of an analysis period. T_N is a factor M times said noise level.

[0099] In the present example, $T_N = -26$ dB; $M=10$. The formula for converting an amplitude A (like 0.005) to dB is $20 \cdot \log_{10}(A)$. This is a common definition. A mathematical property of the log function in there is that $\log_{10}(10 \cdot A)$ equals $\log_{10}(A) + 1$. That is why a 10 times higher amplitude, is 20dB higher.

[0100] For instance, if the average loudness (noise) is Amplitude = 0.005 (=corresponding to -46dB) then $T_N = 10 \cdot 0.005 = 0.05$ (corresponding to = -26dB).

[0101] In the present example, the third parameter T_{min} is a static parameter corresponding to the minimum loudness threshold, here $T_{min} = -32$ dB.

[0102] The present invention is further illustrated by three examples supported respectively by figure 3, figure 4 and figure 5.

[0103] The three examples, the point is to compare the detected cough by using either the method according to the present invention represented and explained on figure 1 and corresponding detail here above, or the method described in Barata et al. ("Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study", J.Med. Internet Res, vol.22, issue 7, 2020).

Comparison example 1:

[0104] In example 1, the experience is based on the following parameters:

- analyzing period = 4 hours during the night
- patient: one patient and one partner (two distinct persons)
- environment: one patient and one partner in a shared bedroom (note: Partner coughs do not influence the adaptive loudness threshold).

[0105] The results of the comparison example 1 is illustrated on table 1 below.

TABLE 1:

	Inferences run	Coughs detected
<i>Adaptive threshold</i>	355	35 / 36
<i>Static threshold -26dB</i>	371	33 / 36
<i>Barata</i>	1040	13 / 36

[0106] An inference is an application of any of the 5 trained NN models. To process one frame fully (to the point of cough detection) would require 5 inferences. However, if the application of the first model would already give a probability below the first reference value, this frame is discarded and thus would only have run one inference.

Adapting loudness threshold (table 1, figure 3):

[0107] In this first example, the adaptive threshold starts at around -32dB and tends to around -24.5dB. The total number of frames processed is only slightly higher than with Barata but almost all coughs are detected instead of around a third with Barata's. Also, the adaptive threshold processes slightly fewer peaks and find slightly more coughs than with a static threshold. Since Barata also always uses all five models, the number of inferences he runs is way higher. Overall, the adaptive threshold is mostly governed by the noise threshold T_N (i.e., second parameter) and the cough amplitude-based threshold (i.e., first parameter).

Cough detection (table 1, figure 3):

[0108] The true number of coughs in this first example is 36, of which the adaptive thresholding detects 35 using 215 processing steps (see figure 1). The number of coughs detected (=35) is significantly higher compared to the number of coughs detected by Barata's (13).

Comparison of example 2:

Example 2:

[0109] In example 2, the experience is based on the following parameters:

- analyzing period = 4 hours during the night
- patient: one patient and one partner (two distinct persons)

- environment: one patient and one partner in a shared bedroom (note: Partner coughs do not influence the adaptive loudness threshold).

[0110] The results of the comparison example 2 is illustrated on table 2 below.

TABLE 2:

	Inferences run	Coughs detected
<i>Adaptive threshold</i>	1583	296 / 444
<i>Static threshold -26dB</i>	1174	204 / 444
<i>Barata</i>	1620	107 / 444

Adapting loudness threshold (table 2, figure 4)

[0111] In this example the adaptive threshold starts at around -32dB and pretty much stays there (except for noisy periods). The total number of frames processed is still higher than with Barata,

Cough detection (table 2, figure 4):

[0112] However, the adaptive threshold is able to detect significantly many more coughs, that would have gone unnoticed with Barata. The static threshold is also much better than Barata, however still far below the adaptive threshold in terms of coughs detected.

Comparison example 3:

[0113] In example 3, the experience is based on the following parameters:

- analyzing period = 1 hour during the day
- patient: One patient
- environment: One patient monitoring during the day, with a very high level of background noise and other loud sounds (possibly multiple distinct persons around).

[0114] The results of the comparison example 3 is illustrated on table 3 below.

TABLE 3:

	Inferences run	Coughs detected
<i>Adaptive threshold</i>	102	0 / 0
<i>Static threshold -26dB</i>	3191	0 / 0
<i>Barata</i>	29570	0 / 0

Adapting loudness threshold (table 3, figure 5)

[0115] The adaptive threshold is mostly governed by the noise threshold and the minimum threshold. The coughs are rather silent and hard to detect. In this noisy night (for example it could be that the patient did not go to sleep, but had a night out) were no coughs occurred the adaptive threshold shows its strength mainly due to noise exclusion. Both with

Barata and with the static threshold thousands of frames would be processed for nothing. The adaptive threshold would process only 102.

Cough detection (table 2, figure 4):

[0116] The true number of coughs in this first example is 444, of which the adaptive thresholding detects 296 using 399 processing steps (preprocessing and neural network). The adaptive threshold is mostly governed by the noise threshold and the minimum threshold. The coughs are rather silent and hard to detect.

[0117] While the embodiments have been described in conjunction with a number of embodiments, it is evident that many alternatives, modifications and variations would be or are apparent to those of ordinary skill in the applicable arts. Accordingly, this disclosure is intended to embrace all such alternatives, modifications, equivalents and variations that are within the scope of this disclosure. This for example particularly the case regarding the different apparatuses which can be used.

Claims

1. Computer-implemented method for detecting a patient's acoustic events, for instance coughs, from audio data of the patient during an analyzing period, the method comprising:

- a recording step to record audio signals from the patient by using a mobile device comprising recording means, and
- a processing step to process the recorded audio signals into audio data by using a processing module embedded in the device,
- a detection step to detect said acoustic events from the audio data,

characterized in that the method further comprises a monitoring step, before the recording step, to select segments of a determined duration of the patient's audio signal to be recorded on the device, said monitoring step comprising:

- applying an adaptive loudness threshold named ALT on the patient's audio signal to identify sounds that exceed said adaptive loudness threshold named relevant sounds,
- said adaptive loudness threshold being based on an initial loudness threshold that is configured to be adjusted continuously during said analyzing period, and
- recording on the device the segments of the patient's audio signal located downstream said identified relevant sounds.

2. Method according to claim 1, wherein the adaptive loudness threshold is adjusted based on at least three parameters defined as:

- a first parameter named T and based on the loudness of the detected acoustic events from the patient;
- a second parameter named T_N and based on a factor M times of the average loudness of the audio signal;
- a third parameter named T_min and based on a determined minimum threshold;

wherein said adaptive loudness threshold ALT corresponds to the maximum value chosen among the first parameter, the second parameter and the third parameter and is defined as

$$ALT = \max (T, T_{\min}, T_N)$$

3. Method according to claim 2, wherein the first parameter T is based on the distribution of the acoustic event detected in the patient, said distribution being adjusted continuously as the acoustic events are detected.

4. Method according to any one of claims 2 to 3, wherein the factor M time of the second parameter T_N is chosen between 0 and about 100, preferably about 2 and about 50, more preferably between about 4 and about 25, in particular between about 8 and about 12.5, for instance between about 9 and about 11, for instance 10.

5. Method according to any one of claims 2 to 4, wherein the minimum threshold of the third parameter is comprised between about -18dB and about -45dB, preferably between about -22dB and about -42dB, more preferably between about -27dB and about -37dB, in particular between about -30dB and about -34dB, for instance between about -31 dB and about -33dB, for instance about -32dB.

6. Method according to any one of claims 1 to 5, wherein the detection step comprises;

- i) Applying a first trained inference model on said recorded segment of audio data of the patient as an input, to compute a probability as an output, said audio data contains at least one acoustic event, said first trained inference model belonging to an ensemble of trained inference models comprising "n" trained inference models, "n" superior to two, to detect said acoustic events; and if the probability is inferior to a first reference value, the acoustic event is rejected, whereas if the probability is superior to the first reference value, step i) is repeated with the next trained inference model from said ensemble comprising "n" trained inference model; and
- ii) the acoustic event is validated if the probability is superior to the first reference value for all the "n" trained models, and if there is a mean probability for all the "n" trained inference model of the ensemble that is above a second reference value;

7. Method according to any one of claims 1 to 6, wherein the first reference value is correlated with the mean probability for all the inference model of the ensemble and the number "n" of trained inference models by using the following formula:

$$\text{First reference value} = 1 - ((1 - \text{second reference value}) \times \text{number "n"})$$

8. Method according to any one of claims 1 to 7, wherein the processing step comprises a pre-processing step that comprises:

- Applying a peak detection filter on said recorded segments to extract a determined number of peaks per recorded segment;
- Processing the extracted peaks of the patient's audio signal to provide the audio data;

9. Method according to any one of claims 1 to 8, wherein the number of peak detectable per recorded segment is adaptive and is determined depending on the processing module specifications, for instance the processing speed or the calculation power of said processing module or acoustic environment.

10. Method according to any one of claims 1 to 9, wherein the method further comprises, before the monitoring step, a status check step comprising:

- determining if the device is in use or not in use based on at least one sensor of the device, and starting the monitoring step if the device is not in use; and/or
- determining if the device has an energy level above a predetermined level, and starting the monitoring step if the energy level is above the predetermined level;

11. Method according to the preceding claim, wherein the status check is based on sensors installed in the device capable of giving information on the position of the device, for instance an accelerometer, and/or on the status of the device, for instance, screen brightness, battery status.

12. Method according to any one of claims 1 to 11, wherein the patient is comprised in a group of "p" patients comprising at least two patients, the method further comprising an assigning step for assigning a detected acoustic event to a patient from the said group, the assigning step comprising:

- Applying a computational tool, for instance a statistical model on the detected acoustic event, said computational tool being configured to assign a detected acoustic event to a predetermined patient;
- Attributing the detected acoustic event to one patient from the group if there is a correlation between said computation tool and the detected acoustic event;

13. Method according to claim 12, wherein

either the computation tool has been trained to assign a detected acoustic event to a type of patient is based on gender, for instance male or female, age for instance child or adult;

or the computation tool has been trained to assign a detected acoustic event to a designated patient based on previously detected acoustic event from said designated patient.

14. Device comprising means for carrying out the method of any one of claims 1 to 13 preferably the device is chosen among a mobile device, for instance a smartphone.

15. Computer program comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method of any one of claims 1 to 13.

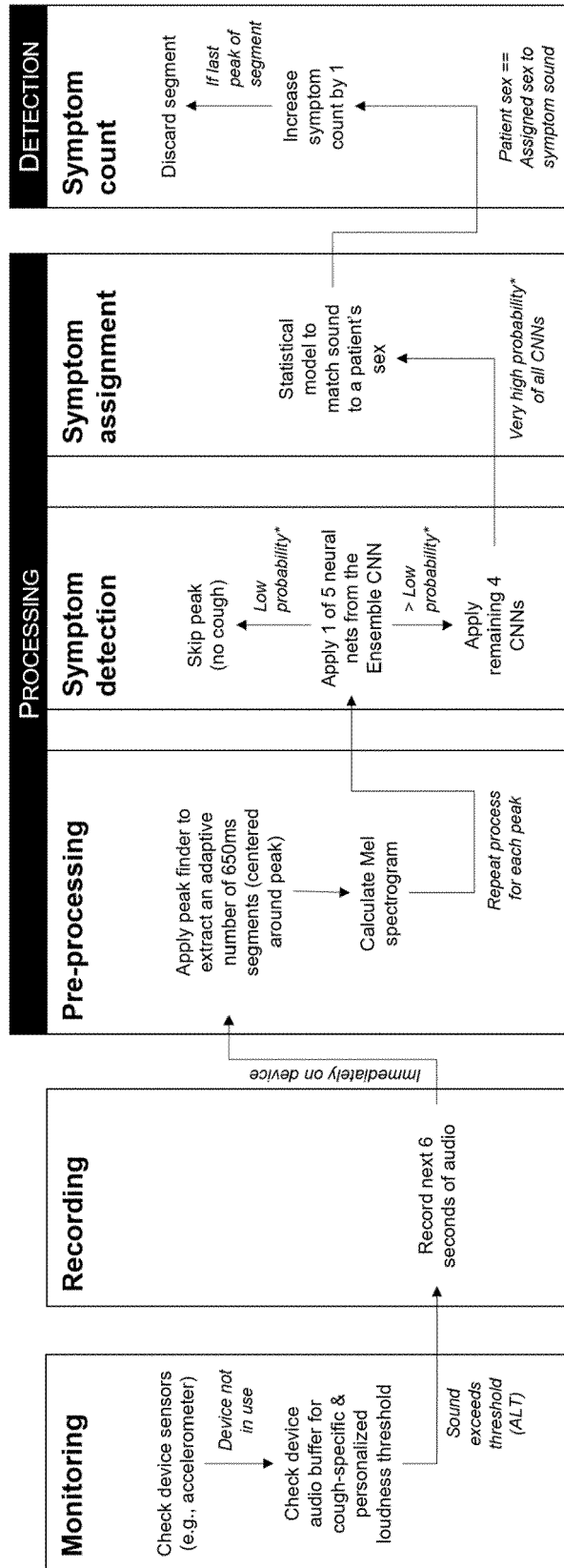


Figure 1

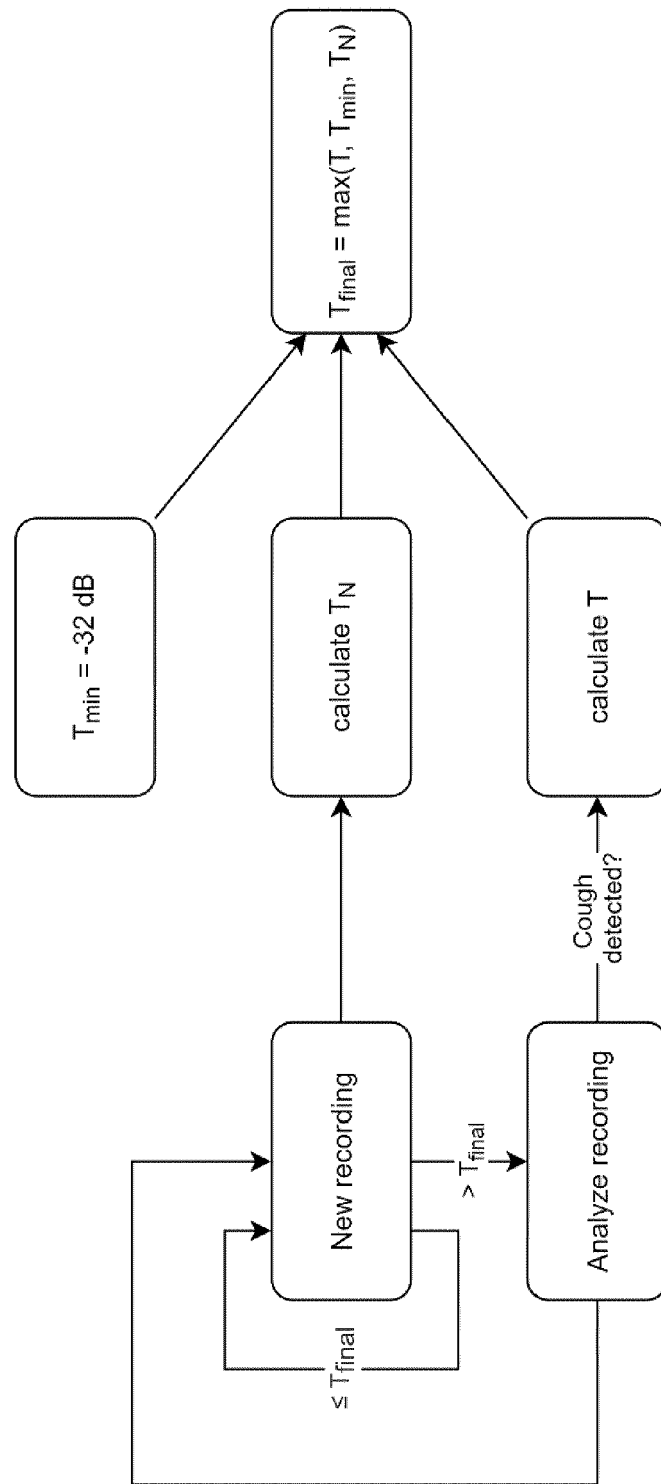


Figure 2

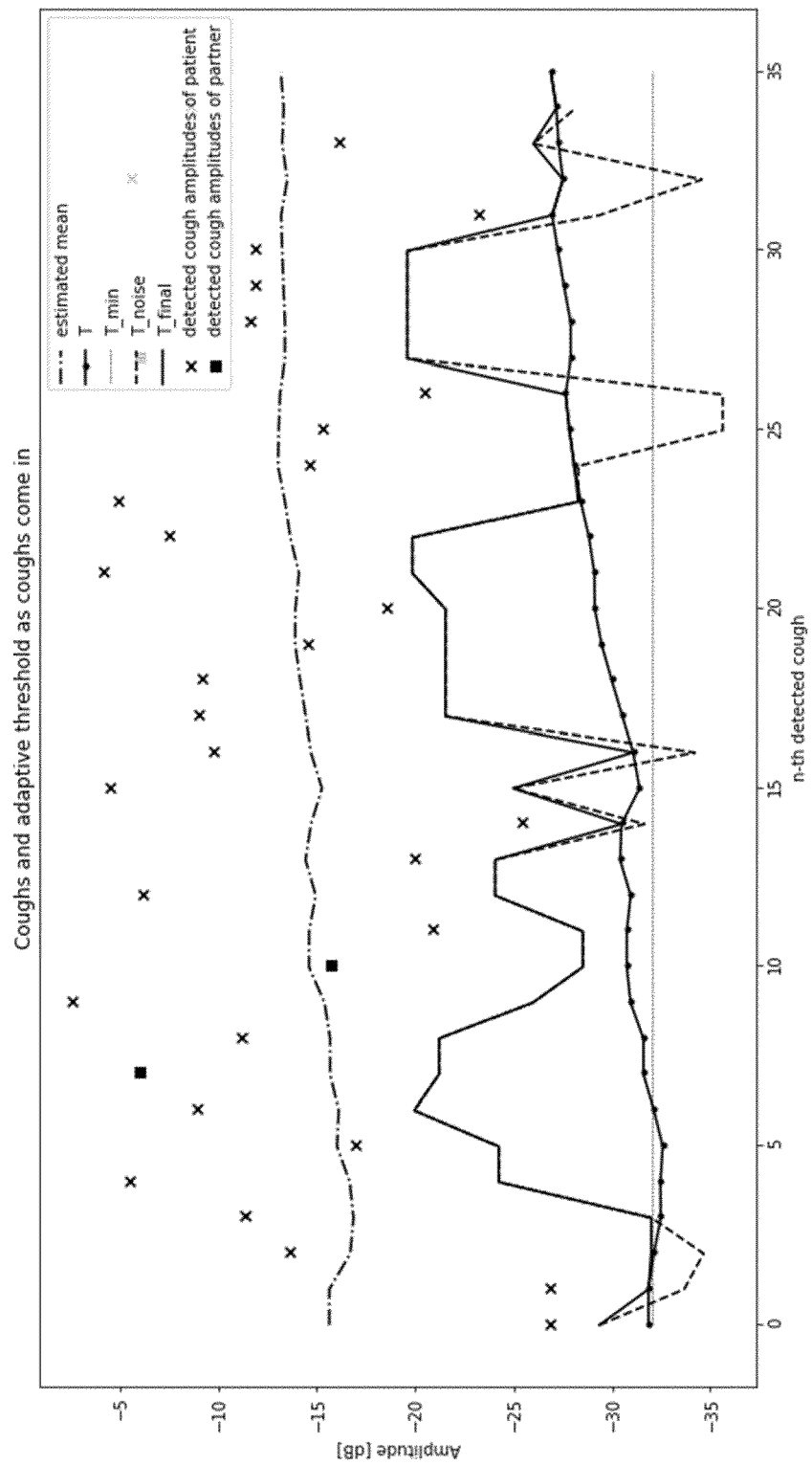


Figure 3

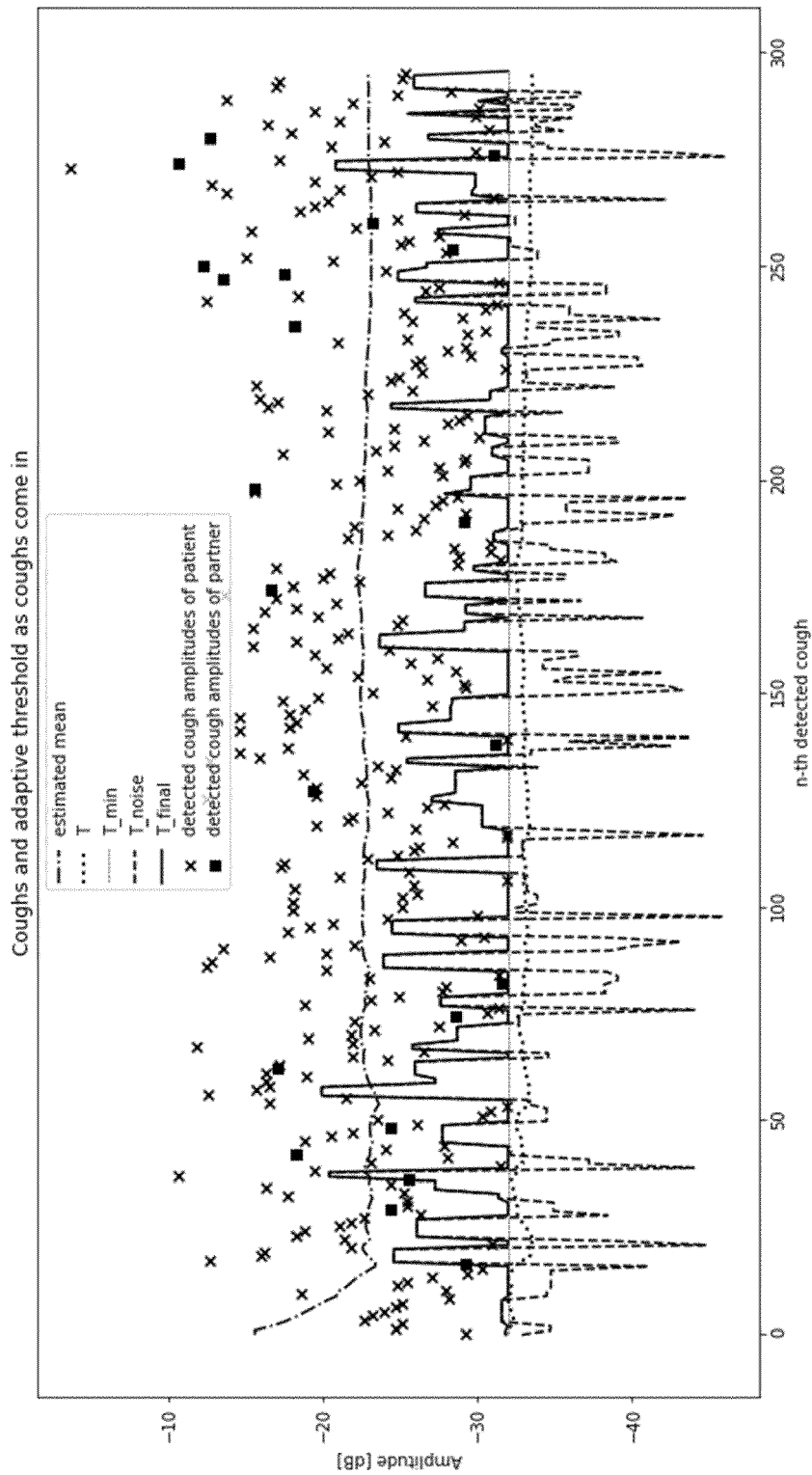


Figure 4

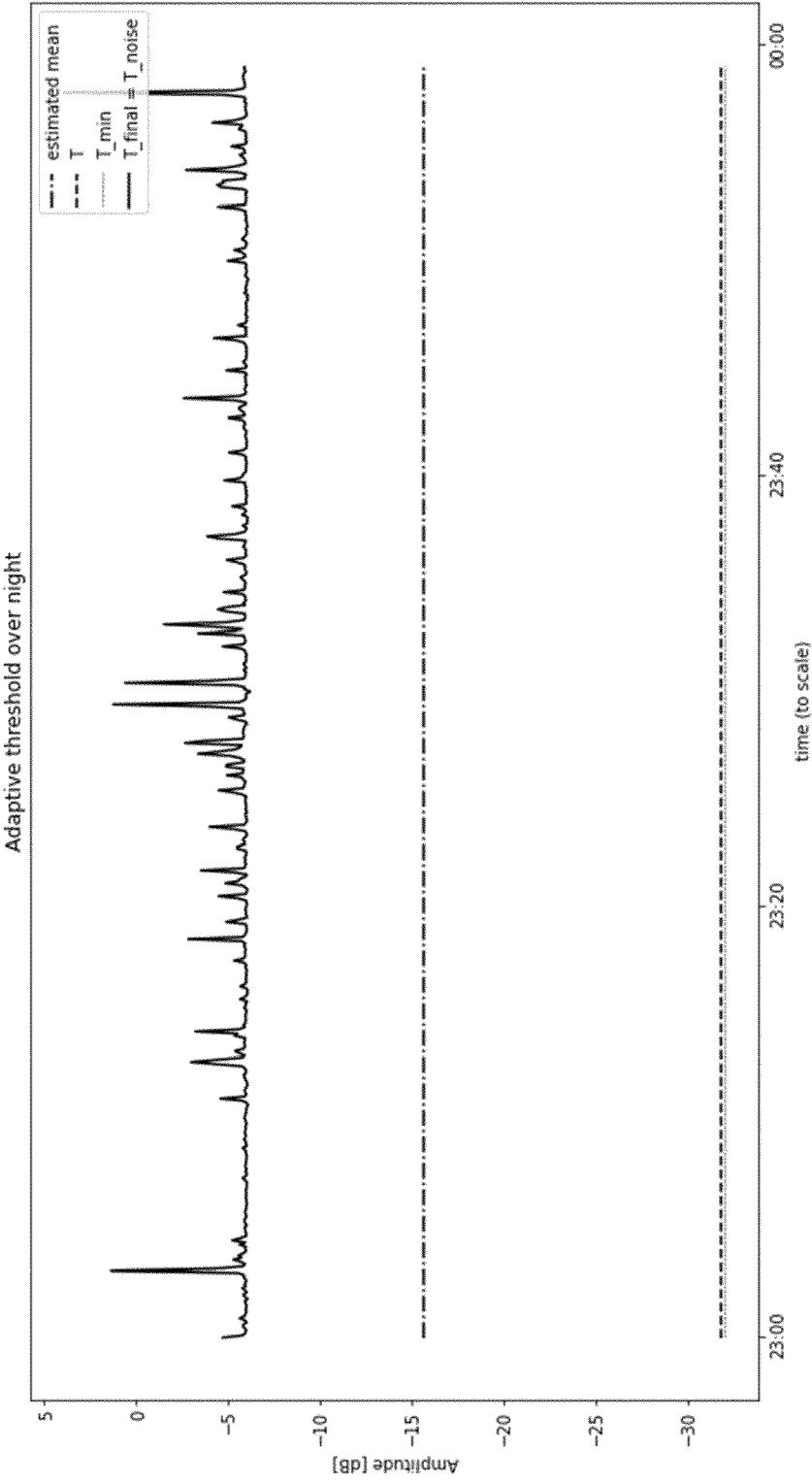


Figure 5



EUROPEAN SEARCH REPORT

Application Number
EP 21 18 3483

5

10

15

20

25

30

35

40

45

50

55

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	US 2015/245788 A1 (SCHMIDT STEVEN P [US]) 3 September 2015 (2015-09-03)	14	INV. A61B7/00
Y	* paragraphs [0052], [0056] - paragraph [0058] *	1,12,13,15	
A	* paragraph [0091] - paragraph [0094] * * figure 5A *	2-11	
Y	----- US 2006/074334 A1 (COYLE MICHAEL [US]) 6 April 2006 (2006-04-06) * paragraph [0050] *	1,12,13,15	TECHNICAL FIELDS SEARCHED (IPC) A61B
Y	----- WO 2006/117780 A2 (GAVRIELY OREN [IL]) 9 November 2006 (2006-11-09) * page 8, line 15 - line 17 *	1,12,13,15	
The present search report has been drawn up for all claims			
Place of search Munich		Date of completion of the search 18 November 2021	Examiner Doyle, Aidan
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	

 1
EPO FORM 1503 03.82 (P04C01)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 21 18 3483

5

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

18-11-2021

10

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015245788 A1	03-09-2015	NONE	
US 2006074334 A1	06-04-2006	US 2006074334 A1 WO 2007001431 A2	06-04-2006 04-01-2007
WO 2006117780 A2	09-11-2006	AU 2006242838 A1 EP 1879501 A2 US 2009216127 A1 US 2012302921 A1 WO 2006117780 A2	09-11-2006 23-01-2008 27-08-2009 29-11-2012 09-11-2006

15

20

25

30

35

40

45

50

55

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **BARATA et al.** Automatic Recognition, Segmentation, and Sex Assignment of Nocturnal Asthmatic Coughs and Cough Epochs in Smartphone Audio Recordings: Observational Field Study. *J.Med. Internet Res*, 2020, vol. 22 (7 **[0006]** **[0062]** **[0086]** **[0103]**)